Division III Women’s Soccer 2022 Data: Correlations and Two-Way Tables

The dataset **d3\_wsoc2022.csv** contains all the game results from the 2022 Division III women’s soccer season. The various variables in the dataset include away\_team, away\_score, home\_team, home\_score and date of the game. This worksheet examines correlation and whether there is an association between the home score and the away score.

**Part 1: Correlation**

1. Before calculating the correlation value between home\_score and away\_score, do you think that the value will be positive or negative? Explain.
2. Calculate the correlation value for home\_score vs. away\_team. Interpret the correlation value in context.
3. Perform a test of association to see if the correlation value that you calculated above is statistically significant. Make sure to state the hypotheses clearly and follow the appropriate steps.
4. Plot home\_score vs. away\_score as a scatterplot. Does there appear to be a relationship between the two variables? If so, describe it. If not, explain the problem.

**Part 2: Chi-Square and Two-Way Table**

1. From the scatterplot in question 4, it is evident that there are a lot of games with many different scores that range from 0-0 all the way to 18-0. Let’s see if there is a significant association between home\_score and away\_score when we group the games by scores and treat the various groups as categorical variables. Using statistical software, create a two-way contingency table that separates the games into 0, 1, 2, and 3+ goals for the home team and the away team.
2. Perform a chi-square test. State the hypotheses clearly and interpret the p-value in context.
3. Find the largest discrepancies between the expected counts and observed counts. Explain how these values confirm that there is a positive or negative correlation between away\_score and home\_score.